Docket No.: 62807-158 **PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of	:	Customer Number: 20277
Makio MIZUNO	:	Confirmation Number:
Serial No.:	:	Group Art Unit:
Filed: January 27, 2004	:	Examiner:
For: STORAGE SYSTEM	:	

CLAIM OF PRIORITY AND
TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT

Mail Stop CPD
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the provisions of 35 U.S.C. 119, Applicant hereby claim the priority of:

Japanese Patent Application No. JP 2003-369811, filed on October 30, 2003.
Japanese Patent Application No. JP 2003-116451, filed on April 22, 2003.

cited in the Declaration of the present application. A certified copy is submitted herewith.

Respectfully submitted,

MCDERMOTT, WILL & EMERY

Keith E. George
Registration No. 34,111

600 13th Street, N.W.
Washington, DC 20005-3096
(202) 756-8000 KEG:gav
Facsimile: (202) 756-8087
Date: January 27, 2004

62807-158
Makio Mizuno
January 27, 2004

日本国特許庁 JAPAN PATENT OFFICE

McDermott, Will & Emery

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: ２００３年１０月３０日

出願番号
Application Number: 特願２００３－３６９８１１

[ST. 10/C]: [ＪＰ２００３－３６９８１１]

出願人
Applicant(s): 株式会社日立製作所

２００４年 １月 ８日

特許庁長官
Commissioner,
Japan Patent Office 今井康夫

出証番号 出証特２００３－３１０９６８３

【書類名】　特許願
【整理番号】　K03013991A
【あて先】　特許庁長官殿
【国際特許分類】　G06F 15/16
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内
　【氏名】　水野　真喜夫
【特許出願人】
　【識別番号】　000005108
　【氏名又は名称】　株式会社　日立製作所
【代理人】
　【識別番号】　100075096
　【弁理士】
　【氏名又は名称】　作田　康夫
【先の出願に基づく優先権主張】
　【出願番号】　特願2003-116451
　【出願日】　平成15年 4月22日
【手数料の表示】
　【予納台帳番号】　013088
　【納付金額】　21,000円
【提出物件の目録】
　【物件名】　特許請求の範囲　1
　【物件名】　明細書　1
　【物件名】　図面　1
　【物件名】　要約書　1
　【包括委任状番号】　9902691

【書類名】特許請求の範囲
【請求項１】
ネットワークに接続されたキャッシュストレージ装置であって、
前記ネットワークにはさらに、記憶媒体上の論理アドレスとデータ長を指定するブロックデータを送受信する1つ以上のクライアント及び１つ以上の記憶装置が接続され、前記クライアントと前記記憶装置の間で送受信する前記データを一時的に蓄積することを特徴とするキャッシュストレージ装置。
【請求項２】
前記クライアントから前記キャッシュストレージ装置に対してライト要求が発行された場合、前記ライト要求が示す前記キャッシュストレージ装置内の領域をロックすることを特徴とする請求項１記載のキャッシュストレージ装置。
【請求項３】
前記クライアントから前記キャッシュストレージ装置に対してライト要求が発行された場合、前記ライト要求が示す前記キャッシュストレージ装置内の領域をロックするとともに前記キャッシュストレージ装置内の領域に対応する前記記憶装置内の領域もロックすることを特徴とする請求項１記載のキャッシュストレージ装置。
【請求項４】
前記ロックをするかしないかを、前記キャッシュストレージ装置内のロック状況を示すロック管理テーブルをもとに判断することを特徴とする請求項２または３に記載のキャッシュストレージ装置。
【請求項５】
前記ロック管理テーブルは、少なくとも前記キャッシュストレージ装置内の領域を判別するインデックス、前記キャッシュストレージ装置内の領域の前記ロックの状況を示すフラグ、前記キャッシュストレージ装置内の領域に対応する前記記憶装置内の領域の前記ロックの状況を示すフラグで構成することを特徴とする請求項４記載のキャッシュストレージ装置。
【請求項６】
前記クライアントから受信した前記ロックの要求に対する許可を前記クライアントへ発行したあと、前記許可に対する応答確認が無かった場合、前記ロックの要求を無効とすることを特徴とする請求項２または３に記載のキャッシュストレージ装置。
【請求項７】
前記キャッシュストレージ装置内の領域に対する処理の要求が前記クライアントから無い場合に、前記キャッシュストレージ装置内の領域に対応する前記記憶装置内の領域に対して前記キャッシュストレージ装置内の領域の内容のライト要求を発行することを特徴とする請求項２または３に記載のキャッシュストレージ装置。
【請求項８】
前記キャッシュストレージ装置内の領域と前記記憶装置内の領域との対応を示すアドレス対応テーブルを備える請求項７記載のキャッシュストレージ装置。
【請求項９】
前記記憶装置へデータを送信するときに、該データに対して暗号化を施すことを特徴とする請求項７記載のキャッシュストレージ装置。
【請求項１０】
前記クライアントから一定時間要求が無かった場合、前記ロックしている前記記憶装置内の領域の解放要求を前記記憶装置に対して発行することを特徴とする請求項２または３に記載のキャッシュストレージ装置。
【請求項１１】
前記クライアントの認証を行って、通信を許可した時点で、前記クライアントに対してアクセスを許可する前記キャッシュストレージ装置内の領域をロックすることを特徴とする請求項１記載のキャッシュストレージ装置。
【請求項１２】

前記クライアントの認証を行って、通信を許可した時点で、前記クライアントに対してアクセスを許可する前記キャッシュストレージ装置内の領域をロックするとともに前記キャッシュストレージ装置内の領域に対応する前記記憶装置内の領域もロックすることを特徴とする請求項１記載のキャッシュストレージ装置。

【請求項１３】

前記キャッシュストレージ装置内の領域に対する処理が前記クライアントから無い場合に、前記キャッシュストレージ装置内の領域に対応する前記記憶装置内の領域に対して前記キャッシュストレージ装置内の領域の内容のライト要求を発行することを特徴とする請求項１１または１２に記載のキャッシュストレージ装置。

【請求項１４】

前記記憶装置へデータを送信するときに、該データに対して暗号化を施すことを特徴とする請求項１３記載のキャッシュストレージ装置

【請求項１５】

請求項１において、前記クライアントからのリード要求を受けた場合前記キャッシュストレージ装置上に該当するデータが存在する場合そのデータを前記クライアントに送信し、存在しない場合は前記記憶装置に該当する前記データを要求し、前記記憶装置から送られてきた前記データを前記クライアントに送信することを特徴とするキャッシュストレージ装置。

【請求項１６】

前記クライアント、前記記憶装置の識別情報を管理する識別情報管理手段が前記ネットワークに接続されているネットワークストレージシステムにおいて、前記識別情報管理手段に登録されている前記記憶装置の識別情報を変更する手段を備えることを特徴とする請求項１記載のキャッシュストレージ装置

【請求項１７】

記憶媒体上の論理アドレスとデータ長を指定するブロックデータを送受信する1つ以上のクライアント及び１つ以上の記憶装置、並びに前記記憶装置の識別情報を管理する識別情報管理手段がネットワークに接続されており、さらに前記クライアントと前記キャッシュストレージ装置との間に前記クライアントの処理を代行する代行処理装置が接続されているネットワークストレージシステムであって、前記代行処理装置が該キャッシュストレージ装置の前記識別情報を前記識別情報管理手段から取得し、前記代行処理装置がその識別情報を元に該キャッシュストレージ装置に対する処理を前記クライアントに代わり実行し、その処理結果を前記クライアントに送信することを特徴とするネットワークストレージシステム。

【請求項１８】

記憶媒体上の論理アドレスとデータ長を指定するブロックデータを送受信する1つ以上のクライアント、及び１つ以上の記憶装置、並びにキャッシュストレージ装置を備え、さらに前記クライアント、前記記憶装置、前記キャッシュストレージ装置とを接続する結合装置を備え、

前期結合装置が、前記クライアントから前記記憶装置への要求を一旦前記キャッシュストレージ装置へ送信し、

前記伽シュストレージ装置がその要求を処理し、その処理結果を前記クライアントへ送信した後で、

前記キャッシュストレージ装置から記憶装置へ前記要求を発行することを特徴とするネットワークストレージシステム。

【書類名】明細書
【発明の名称】ストレージ装置及びネットワークシステム
【技術分野】
【０００１】
本発明は、サーバとストレージとの間のネットワークを介したデータ転送においてレスポンス性能の改善、さらに障害発生時のオーバヘッドを削減する技術に関する。
【背景技術】
【０００２】
記憶装置（以下「ストレージ」とも言う）の接続形態は、計算機（以下「サーバ」とも称する）直結型のDirect Attached Storage(DAS)からネットワーク接続型のStorage Area Network(SAN)が主流となっている。SANを実現する伝送方式としてファイバチャネルを用いたFC-SANが一般的である。
【０００３】
一方、ファイバチャネルに対して転送性能で遅れを取っていたイーサネット（登録商標）が、ネットワーク技術の進歩に伴い高速化されてSANへ適用されつつある。これをFC-SANと区別するためにIP-SANと呼ぶ。IP-SANを実現する手段としていくつか候補があるが、iSCSI(internet SCSI)が最有力と言われている。
【０００４】
信頼性が低いイーサネット（登録商標）に基づいたiSCSIでは、データ転送の信頼性を確保するためにTCP/IPプロトコルが用いられる。しかし、TCP/IPを採用すると信頼性が確保される反面、データ転送に係るオーバヘッドが大きくなるという問題がある。
【０００５】
例えば、TCPはコネクション型であり、TCP層を介して送られて来るデータの順序補償、誤り訂正、障害発生時の再送処理が行われる。特に障害発生時の再送処理のオーバヘッドはサーバとストレージ間の接続距離に比例して大きくなり性能に影響を及ぼす。
【０００６】
このTCPにおける再送処理の問題を解決する手段として、サーバとストレージの間にデータを一時保存（以下「キャッシュ」）する装置（以下「キャッシュデバイス」）を配置し、サーバからの要求をそのキャッシュデバイスでキャッシングする方法がある。再送処理が発生した場合は、サーバからではなくキャッシュデバイスからデータ等を再送することにより上記オーバヘッドが削減される。
【０００７】
上述のキャッシュデバイスの考え方は、Webアクセスの技術においては、「Webキャッシュ」として採用されている。具体的には、Webにおいて、サーバから取得したページを一旦ローカルのキャッシュデバイスにキャッシュするのが一般的である。
【０００８】
このようなWebキャッシュでは、サーバ側でページを絶えず更新していることを考慮すると、ローカルにキャッシングしたページとサーバで管理するページとで鮮度を保証（データの一致を保証）する必要がある。
【０００９】
Webアクセスを実現するプロトコルHTTP(HyperText Transport Protocol)では鮮度を保証する手段として、Webサーバへの要求に対するレスポンスヘッダに定義されているAge等の情報に基づいて、ローカルにキャッシュされたデータを使用可能かどうかをローカルで判定する方法がある。ただし、ローカルでキャッシュ可能なデータはリード（サーバからの読み込み）系のコマンドに対応するものに限られている。
【００１０】
更に、ファイルシステムにおけるデータのキャッシュ方法に関しては、特許文献１及び特許文献２に開示されている。
特許文献１では、ファイルサーバが接続されたネットワークに、クライアントがキャッシュサーバ（キャッシュデバイスの一種）を介して接続されたシステムにおいて、アクセ

ス頻度の高いファイルをキャッシュサーバ上に先読みしておくことで高速なファイルアクセスを実現する技術が開示されている。

【００１１】

特許文献２では、サーバとキャッシュサーバとクライアントがネットワークを介して接続されたシステムにおいて、データを要求したクライアントへのデータ配信の負荷が小さく、通信時間の遅延が小さいキャッシュサーバのみを介してデータを配信することで、無駄なデータが他のキャッシュサーバに蓄積されることを防ぐ技術が開示されている。

【００１２】

【特許文献１】特開平１１－２４９８１公報

【００１３】

【特許文献２】特開２００１－２９０７８７公報

【発明の開示】

【発明が解決しようとする課題】

【００１４】

上述した公知技術はファイルデータを対象としており、かつリードデータのみがキャッシュデバイスにキャッシングされる。尚、ファイルデータとはファイルシステムがアクセス可能な単位で構成されたデータを指す。

【００１５】

しかし、トランザクション性能が要求されるアプリケーションではブロック単位でのデータのアクセスが望ましい。これは、ファイルデータへのアクセスは最終的にブロック単位でのアクセスに変換されるため、その分のオーバヘッドがトランザクション性能に影響を与えるためである。

【００１６】

又、例えばデータベースアプリケーションではリードアクセスだけはなくライトアクセスも行われる。

ライトアクセスで問題になるのは、そのキャッシングされたデータに対して複数のクライアントがアクセスする場合である。複数のクライアントから同一のデータ格納領域にライトアクセスのリクエストが発行された場合、キャッシュデバイスでその双方のリクエストが競合し、双方がデータの書き込み又は書き換えをしようとするためデータの保証が出来ない。このような点に関しては、上述の公知技術では言及がされていない。

【００１７】

さらに公知技術の場合、通信プロトコルによって決められているアルゴリズムにしたがって送信元、受信元の装置間のデータ通信経路が決定される。したがって、ネットワーク上にキャッシュデバイスを用意しても、必ずしもキャッシュデバイスを通過するデータ通信経路にならない場合があり、キャッシュデバイスを採用した利益を享受することが出来ないという問題がある。

【００１８】

一般的に、データ通信経路を決定するアルゴリズムでは、キャッシュデバイスの存在は考慮されず、データ転送に使用されるネットワークの数（Hop数）やコストに基づいて決定される。したがって、公知技術ではデータ転送において必ずキャッシュデバイスを使用するという保証をすることができない。

【００１９】

本発明の目的は、サーバとストレージ間のデータ転送の転送効率を上げることにある。

本発明の別の目的は、キャッシュデバイスにおいてクライアントが複数存在する場合にリクエストの競合を防止することにある。

本発明のさらに別の目的は、ネットワーク上のキャッシュデバイスにクライアントが通信できるようにすることにある。

【００２０】

本発明のさらに別の目的は、サーバ、ストレージ及びキャッシュデバイス間のデータ転送のセキュリティを確保することにある。

【課題を解決するための手段】
【００２１】
上記問題を解決するために、クライアントとストレージの間にネットワークに接続可能でブロック単位のデータを一時的に蓄積するキャッシュデバイスを設ける。そして、ストレージと通信を行おうとするクライアントに対し、ネットワークに存在する管理端末、例えばネームサーバ等を利用してクライアントにストレージの代わりにキャッシュデバイスを指定する。
【発明の効果】
【００２２】
本発明によれば、サーバとストレージの間で、ブロック単位のデータのキャッシングによるレスポンス性能、トランザクション性能の向上、およびクライアント間のデータの一貫性、新鮮さ、セキュリティを保証したデータ通信を行うことが可能となる。
【発明を実施するための最良の形態】
【００２３】
以下本発明の実施形態について説明する。
【実施例１】
【００２４】
以下、本発明に係わるキャッシュデバイス（以下、キャッシュデバイスがストレージの場合について説明し、そのストレージを「キャッシュストレージ装置」又は「キャッシュストレージ」とも称する）の実施例１を図面に示しさらに詳細に説明する。
図１は、本実施例のキャッシュストレージ装置を含むネットワークシステムの全体構成を示している。
【００２５】
本実施形態のネットワークシステムは、複数のネットワーク120、135、ネットワークに接続されるクライアント105、キャッシュストレージ125、ネームサービス110、記憶装置130及び各ネットワーク相互を接続するネットワーク結合装置140を有する。
【００２６】
クライアント105は、ストレージターゲットに対してリクエストを発行する計算機である。ここで、ストレージターゲットとはクライアント105と通信するストレージ130を指す。
ネームサービス110は、ネームサービスを提供する計算機であり、TCP/IPネットワークにおいてはDNS(Domain Name System)、iSCSIにおいてはiSNS(Internet Storage Name Service)サーバ、SLP DA(Service Location Protocol-Directory Agent)などに該当する。本図では、ネームサービス110aがDNSサーバに対応し、ネームサービス110ｂがiSNSサーバに対応するとする。
【００２７】
ネームサービス110等は、独立したネットワーク毎に存在する。図1では、2つのネットワークが存在し各ネットワークに1つのネームサービスが存在する。
通常、耐障害性を高める目的でネームサービスは冗長構成にするが図1では省略する。
【００２８】
キャッシュストレージ125は、クライアント105が記憶装置130と行う通信においてクライアント105から送られるデータを一時蓄積する。
【００２９】
記憶装置130は、ディスクドライブ等のストレージデバイスを有する装置、例えばディスクアレイ装置である。
ネットワーク135には、クライアント105c、キャッシュストレージ125b、記憶装置130、ネームサービス110bが接続されている。
ネットワーク120には、クライアント105a、105b、ネームサービス110aが接続される。ネットワーク120には、例えばLAN(Local Area Network)などが採用される。
【００３０】

ここで、ネットワーク120をネットワーク1、ネットワーク135をネットワーク2と区別しそれぞれ独立した異なるネットワークと仮定する

ネットワーク結合装置140は、異なるネットワークを接続するための装置である。

【００３１】

以下、簡単にネームサービスについて説明する。

DNSとは、ネットワークに接続された各クライアントに与えられる識別子（以下「ホスト名」）から対応するIPアドレスを取得する仕組みであり、DNSサーバがその対応をデータベースで管理する。

【００３２】

図１で、例えばクライアント105aのホスト名がabc、IPアドレスが192.168.0.1、クライアント105ｂのホスト名がdef、IPアドレスが192.168.0.10とする。

クライアント105aがクライアント105ｂと通信する場合、クライアント105aは、クライアント105ｂのホスト名を用いてネームサービス110aにクライアント105ｂのIPアドレスを問い合わせる。

【００３３】

問い合わせを受けたネームサービス110aは、ネームサービス110aが管理しているデータベースからクライアント105ｂのホスト名に対応するIPアドレスをクライアント105aへ返す。

クライアント105aは、ネームサービス110aから受け取ったIPアドレスで初めて他のクライアント105ｂとの通信が可能となる。

【００３４】

一方、iSNSは、IPネットワーク上のiSCSI対応の記憶装置130（以下「iSCSIストレージ」）とファイバチャネルに対応する記憶装置130（以下「ファイバチャネルストレージ」）の管理フレームワークであり、iSNSサーバがネットワーク上のクライアントとストレージを管理する。

【００３５】

クライアントやストレージの識別には、iSCSIストレージの場合iSCSIネーム、ファイバチャネルストレージの場合WWPN(World Wide Pote Name)が使用される。

以下、簡単にiSNSのネームサービスの手順について図１を用いて説明する。図１において、ネームサービス110bを介してクライアント105cと記憶装置130aが通信する場合を考える。

【００３６】

ここで、記憶装置130aがネットワーク2に接続され、ネームサービス110bに記憶装置130aの情報が既に登録されているものとする。

クライアント105cがネットワーク2に接続されると、クライアント105cは、まず自分自身の情報をネームサービス110bへ登録する。

これにより、ネットワーク2に存在するストレージ130aにクライアント105cが同一ネットワーク上に存在するという通知がネームサービス110bから送られる。

【００３７】

そして、クライアント105cは、記憶装置130aを検出するための要求（以下「クエリー」）をネームサービス110bへ送信し、その応答として、記憶装置130aの情報（iSCSIネーム、IPアドレス、ポート番号）を取得する。

以上を経て、クライアント105cはiSCSIストレージデバイス130へ通信可能となる。

【００３８】

図２は、キャッシュストレージ125及び記憶装置130の構成例を示す図である。

キャッシュストレージ125は、ディスク制御装置205及びディスク駆動装置235とを有する。

ディスク制御装置205は、入出力パス206、チャネル制御部210、キャッシュメモリ制御部215、共有メモリ制御部220及びディスク制御部230を有する。

【００３９】

入出力パス206は、ネットワーク2とチャネル制御部210とを接続する通信線である。

チャネル制御部210は、ネットワークインタフェースを持ち、クライアントとのユーザデータの送受信、ディスク制御装置205内部の制御情報などの共有データへのアクセスを制御する。チャネル制御部210は、ディスク制御装置205内に複数存在する。

【００４０】

キャッシュメモリ制御部215は、クライアント、及びディスク駆動装置235にあるユーザデータを一時的に格納するキャッシュメモリを有し、チャネル制御部210、またはディスク制御部230からのキャッシュメモリへのアクセスを制御する。

【００４１】

共有メモリ制御部220は、ディスク制御装置205内部での通信に関する制御情報を格納する共有メモリを備え、チャネル制御部210、及びディスク制御部230からの共有メモリアクセスを制御する。

共有メモリ制御部220は、キャッシュストレージ125内のデバイスのロック状況を示すロック管理テーブル225及びアドレス対応テーブル226を有する。

【００４２】

ここで「デバイス」とは、記憶装置130やキャッシュストレージ125が有する論理的又は物理的な記憶媒体を指し、例えばディスクドライブであったり、これらのディスクドライブから構成される論理ユニットであったりする。以下では、ディスクドライブをデバイスとして取り扱うものとする。

【００４３】

又、「ロック」とは、あるデバイスがある装置によって専有されており、他の装置が使用できない状態を指し、「ロックする」とは、あるデバイスを他の装置が使用できない状態にすることを指す。

【００４４】

ロック管理テーブル225には、どのデバイスがロックされているかを示す情報が登録される。

アドレス対応テーブル226には、キャッシュストレージ125内のデバイスとそれに対応する記憶装置130内のデバイスとの対応付けを示す情報が格納される。

【００４５】

具体的には、クライアント105からのユーザデータの一時蓄積場所であるキャッシュメモリ、もしくはディスク駆動装置235に存在するデータを記憶装置130へ格納するときに指定する記憶装置130内のデバイスのアドレスとキャッシュストレージ125内デバイスのアドレスとを対応付けてアドレス対応テーブル226で管理する。

【００４６】

ディスク制御部230は、ディスク駆動装置235との通信制御、及びキャッシュメモリ・共有メモリへのアクセスなどを行う。チャネル制御部210やディスク制御部230の数に制限はない。

【００４７】

チャネル制御部210、キャッシュメモリ制御部215及び共有メモリ制御部220との間の接続や、ディスク制御部230、キャッシュメモリ制御部215及び共有メモリ制御部220との間の接続はバス接続、またはスター接続などの形態をとる。

ディスク駆動装置235は、複数のディスクドライブ240を有し、ユーザが使用するデータ（以下「ユーザデータ」とも称する）などを格納する。

【００４８】

ここで、キャッシュストレージ125は、ディスク駆動装置235のディスクドライブ240に対して1つ以上のiSCSIネームを割り当てる。尚、キャッシュストレージ125は、他のデバイスに対してiSCSIネームを割り当てても良い。iSCSIネームとはiSCSIプロトコルを認識する個々の機器を識別するための情報である。

図３は、チャネル制御部210の構成例を示す図である。チャネル制御部210は、プロトコル制御部305、チャネル制御プロセッサ310、データ転送制御部320及び共有データ制御部3

25を有する。

プロトコル制御部305は、入出力パス206を介してネットワークパケットを受信する。受信したネットワークパケットにiSCSIパケットが含まれる場合、プロトコル制御部305はさらにその中からiSCSIヘッダ、SCSIコマンド、データなどを取り出しチャネル制御プロセッサ310へ渡す。iSCSIパケット以外のパケットには適切な処理（例えばICMP（Internet Control Message Protocol）リクエストであればICMPリプライを発行）が施される。

【００４９】

また、プロトコル制御部305は、SCSIコマンドの処理結果をチャネル制御プロセッサ310から受け取りiSCSIパケットを生成する。そしてプロトコル制御部305は、iSCSIパケットをネットワークパケットにカプセル化し入出力パス206を通じてネットワーク2へ送出する。

【００５０】

チャネル制御プロセッサ310は、プロトコル制御部305からのSCSIコマンド、データなどの受信、解析を行い、解析したリクエストの内容に従い、ディスク制御装置内部に指示を与えるプロセッサである。

【００５１】

データ転送制御部320は、チャネル制御プロセッサ305からの指示により、クライアントのユーザデータ転送、及びディスク駆動装置にあるユーザデータを読み出す。

共有データ制御部325は、制御情報等の共有データが格納されている共有メモリへのアクセスを制御する。

【００５２】

図１３は、ネットワークパケットとiSCSIパケットの関係の一例を示す図である。ネットワークパケット1701は、Etherヘッダ1705、IPヘッダ1710、TCPデータグラム1715から構成される。Etherヘッダ1705はデータリンク層に関連する制御情報（MACアドレス等）とデータを含む。IPヘッダ1710はIP層に関連する制御情報（IPアドレス等）とデータを含む。TCPデータグラム1715はTCPヘッダ1725とTCPデータ1730で構成する。TCPヘッダ1725はTCP層に関連する制御情報（ポート番号、シーケンス番号等）を含む。TCPデータ1730にはiSCSIパケットを含む。iSCSIパケットはiSCSIに必要な制御情報とSCSIコマンド、データを含む。

【００５３】

図４は、ロック管理テーブル225の構成例を示す図である。

ロック管理テーブル225は、複数のフィールド（405、410）を持つエントリを、iSCSIネーム対応に複数有する。

フィールド405はiSCSIネームを登録するフィールドである。ここに登録されるiSCSIネームは、キャッシュストレージ内のデバイスを識別するiSCSIネームである。

【００５４】

フィールド410はロックステータスを登録するフィールドである。ロックステータスとは、フィールド405に登録されたiSCSIネームに対応するキャッシュストレージ125内のデバイス及び記憶装置130内のデバイスのロック状況を示す情報である。

【００５５】

フィールド410は更にサブフィールド420及び425を有する。フィールド420はキャッシュロックフラグを登録するフィールドである。キャッシュロックフラグは、"OFF"の場合iSCSIネーム405に対応するキャッシュストレージ内のデバイスがロックされていない状態、"ON"の場合はロックされている状態を示す情報である。

【００５６】

フィールド425は記憶装置ロックフラグを登録するフィールドである。記憶装置ロックフラグは、"OFF"の場合はiSCSIネームで表されるキャッシュストレージ125内デバイスに対応する記憶装置130内のデバイスがロックされていない状態、"ON"の場合はロックされている状態を示す情報である。

このテーブルを上記のように利用することでキャッシュストレージ、及び記憶装置のロ

ック状況を把握することが可能となりアクセスが競合してもデータを保証できる。

図１１は、アドレス対応テーブル226の構成例を示す図である。アドレス対応テーブル226は、複数のフィールド（1305、1310）を有するエントリを、キャッシュストレージが有するデバイス毎に有する。

フィールド1305はデバイス名が格納されるフィールドである。デバイス名とは、キャッシュストレージ125内のデバイスに割り当てられたiSCSIネームに対応する。

【００５７】

フィールド1310は記憶装置アドレスが登録されるフィールドである。記憶装置アドレスとは、フィールド1305に登録されたデバイス名に対応する記憶装置130内のデバイスに与えられたiSCSIネームである。

【００５８】

キャッシュストレージ125が自装置内のデータを記憶装置130へ格納するときにまず記憶装置130内のデバイスに対しログインする。

そのログインするデバイスのiSCSIネームを指定するときにアドレス対応テーブルを参照しログインパラメータとして記憶装置アドレスを使う。このテーブルを上記のように利用することで記憶装置130に対するログインに必要な情報を容易に参照することが出来る。

【００５９】

図５は、実施例１におけるクライアント105毎に記憶装置130が有するデバイスが割り当てられれている時のクライアント105とキャッシュストレージ125との通信手順とキャッシュストレージ125と記憶装置130との通信手順を示すフローチャートである。尚、ここでの通信はライト処理とする。

【００６０】

図５では、クライアント105毎にあらかじめデバイスが割り当てられているためキャッシュストレージ125でリクエストが競合しない。従って、デバイスをロックする処理が不要となる。

まず、クライアント105は、割り当てられたキャッシュストレージ125内のデバイスに対してログインの要求を発行する。この時、本来ならば記憶装置130に到達するログインをキャッシュストレージ125へと導く必要がある。

【００６１】

そのため、本実施形態では、最初にクライアント105がログインするデバイスの情報を取得するときに、ネームサービス110が記憶装置130ではなくキャッシュストレージ125の情報を送信する。ここでキャッシュストレージ125の情報とは、例えばIPアドレス、ポート番号、iSCSIネーム又はこれらの組み合わせである。

【００６２】

キャッシュストレージ125の情報は、キャッシュストレージ125とネームサービス110との間であらかじめ通信して決める。通信の方法としては、例えばTCPのデータで通信する方法や、iSNSのVender Specific Messageでお互いに取り決めたメッセージを通信する方法がある。このようにして、ネームサービス110は、クライアント105からの情報取得要求のときに記憶装置130の情報ではなくキャッシュストレージ125の情報をクライアント105に提供する（ステップ505）。

【００６３】

ログイン要求を受け取ったキャッシュストレージ125は、ログインを許可するステータスと共にログイン応答メッセージをクライアント105に送信する。この時点で、クライアント105とキャッシュストレージ125内のデバイスとのセッションが確立された状態となる（ステップ510）。

【００６４】

その後、クライアント105はキャッシュストレージ125に対してリクエスト、ここではライト要求を発行する（ステップ515）。

キャッシュストレージ125は、そのリクエストを処理（データの書き込み）し、ステー

タスを含むメッセージをクライアント105に送信する。ステータスとは、そのリクエスト処理が正常終了したかどうか、異常終了した場合その要因を示す情報である（ステップ520）。

【００６５】

クライアント105とキャッシュストレージ125は必要に応じてステップ515、520の処理を繰り返す。そして、クライアント105は最後のリクエストを発行し（ステップ525）、キャッシュストレージ125からそのリクエストに対するメッセージを受ける（ステップ530）とクライアント105側の処理を終了する（ステップ535）。

【００６６】

クライアント105との通信を終えた後、キャッシュストレージ125は、記憶装置130へのデータ格納処理に入る。尚、記憶装置130へのデータ格納処理の実行は、クライアント105との通信終了直後でも、適当なタイミングに行っても構わない。

【００６７】

データ格納処理に入ると、まずキャッシュストレージ125は、データ格納先の記憶装置130へログインし、データ格納を指示するライトリクエストを発行する（ステップ540）。

記憶装置130はそのリクエストに従い、所定の領域にキャッシュストレージ125から転送されたユーザデータを格納する（ステップ545）。

その後、記憶装置130はリクエストに対する応答をキャッシュストレージ125へ送信し処理を終了する（ステップ550）。

【００６８】

尚、キャッシュストレージ125がクライアント105からリード要求を受けた際に、その要求されたユーザデータがキャッシュストレージ125に存在する場合、キャッシュストレージ125はそのユーザデータをクライアント105へ送信する。存在しなければ、キャッシュストレージ125は記憶装置130からユーザデータを読み出しクライアント105へ送信する。

【００６９】

記憶装置130からユーザデータを読み出した場合、キャッシュストレージ125がクライアントにそのユーザデータを送信すると共にキャッシュストレージ125上にそのユーザデータを保存し、次回同一要求を受けた場合にそのキャッシュしたユーザデータを使用してもよい。

【００７０】

以上の処理により、従来記憶装置130まで通信するところをクライアント105により近い場所での通信ですむため、クライアント105から見ると応答時間が短縮されレスポンス性能、トランザクション性能の改善につながる。

【００７１】

図１２は、実施例１における伝送路障害が発生した場合の転送手順を示す図である。

なお、図１２では図５のステップ540以降の処理手順を示すが、図５の他のステップで伝送路の障害が発生しても同様の手順で処理される。

【００７２】

ステップ540では、キャッシュストレージ125は、キャッシュストレージ125に格納されたユーザデータを記憶装置130へ格納する処理を指示するリクエストを記憶装置130へ発行する。この際、伝送路で障害が発生したとする（ステップ1405）。

【００７３】

ここで、キャッシュストレージ125は、要求したリクエストに対する応答を受信できる状態になっている。ところが、送ったリクエストが記憶装置130へ到達する前に障害が発生したため、記憶装置130はこのリクエストを処理できない。よって記憶装置130はリクエストに対する応答が出来ない。

【００７４】

キャッシュストレージ125は、応答が無いためタイムアウトを検出し再送処理を開始する。

このような動作により、従来であればクライアント105まで影響が及ぶ再送処理（1415

）が、記憶装置130に近い位置にあるキャッシュストレージ125までの処理（1410）とすることが出来るため、再送処理の負荷軽減につながる。

【００７５】

上述の処理により、クライアント105では、トランザクション性能が改善される。特に、信頼性の低いネットワークであればより効果が大きくなる。

【００７６】

実施例1によれば、ブロック単位のデータ通信において、クライアントと記憶装置間の距離が離れている場合にその間にキャッシュストレージを設置し、クライアントからのブロック単位のデータをキャッシングすることによりレスポンス性能を改善させると共に再送処理の軽減によるトランザクション性能も改善させることが出来る。

【００７７】

以下、本実施形態においてロック制御を行う場合について、図６から図９に示しさらに詳細に説明する。尚、以下特に説明のない部分はロック制御を行わない場合と同じとする。

図６は、ロック制御におけるクライアント105、キャッシュストレージ125及び記憶装置130間の通信手順を示した図である。以下、通信はライト処理として説明する。

【００７８】

ここでは、クライアント105がキャッシュストレージ125のデバイスをロック制御する。つまり、クライアント105からのリクエストがキャッシュストレージ125で競合する場合である。

【００７９】

ここでは、ロック制御をReserve、ReleaseというSCSIコマンドで実現する。

Reserveコマンドとは、領域全体を特定のデバイスのために予約して排他的に占有できるように指定するコマンドである。

ReleaseコマンドはReserveコマンドで排他的に占有している領域全体を解放するコマンドである。

【００８０】

なお、以下の説明では、ログイン処理は既に完了済みとし説明を省略する。

また、キャッシュストレージ125及び記憶装置130のロック管理テーブルの内容の変遷を図７及び図８に示す。尚、ログイン処理が終了した時点でのロック管理テーブルの内容は、それぞれ図７（１）、図８（１）の状態とする。

【００８１】

クライアント105は、キャッシュストレージ125に対してリクエストを発行する前に該デバイスのロック要求を発行する。これにはReserveコマンドが用いられる（ステップ605）。

【００８２】

キャッシュストレージ125は、Reserveコマンドを受けるとキャッシュストレージ125上のReserveコマンドで指定された領域（デバイス全体又はデバイスの一部）のロック状態を確認する。

キャッシュストレージ125は、指定された領域がロックされていなければその領域をロックし、ステータス"Good"をクライアント105へ送信する。

【００８３】

指定された領域が既にロックされていれば、キャッシュストレージ125はステータス"Reservation conflict"をクライアント105へ送信する。

このとき、キャッシュストレージ125がロック要求を発行したクライアント105の状況を確認するためにロックOKを示すコマンドを発行する（ステップ610）。

【００８４】

ロックOKのコマンドを受けたクライアント105は、その確認応答のためにロックAcknowledgeをキャッシュストレージ125に送信する。これは、他のクライアント105からロック要求を受けたときにデットロックを生じさせないために行われる（ステップ625）。

【００８５】
その後、キャッシュストレージ125は、指定された領域に対応するロック管理テーブルの内容を更新（ここでは、図７の（１）から（２）の内容へ更新）する（ステップ620）。
ロックAcknowledgeを送信したクライアント105は、キャッシュストレージ125に対してI/Oリクエストを発行する（ステップ630）。
キャッシュストレージ125はそのリクエストを処理しステータスを含む応答をクライアント105へ送信する（ステップ640）。
【００８６】
この間、他のクライアント105から同一領域に対するロック要求(ステップ645)もしくはリクエストを受けて（ステップ650）も、キャッシュストレージは"Reservation confilct"ステータスをロック要求等を送信したクライアント105に送信しそれら要求を拒絶する。
【００８７】
必要に応じてステップ630、640を繰り返し一連の処理が終了すると、クライアント105はキャッシュストレージ125に対してロック解除要求を発行する。このとき、ロック解除要求としてReleaseコマンドが使用される（ステップ655）。
キャッシュストレージ125は、Releaseコマンドで指定される領域を開放し、ステータス"Good"をクライアント105へ送信する。このとき、Reserve時と同様にロック解除要求を発行したクライアント105の状況を確認するためにロック解除OKを示すコマンドを発行する（ステップ660）。
【００８８】
ロック解除OKを示すコマンドを受けたクライアント105は、その確認応答のために解除Acknowledgeをキャッシュストレージ125に送信する（ステップ665）。
解除Acknowledgeを受信したキャッシュストレージ125は、Releaseコマンドで指定される領域に対応するロック管理テーブルを更新（ここでは図７の（２）から（３）の内容へ更新）する（ステップ657）。
【００８９】
この後、キャッシュストレージ125は、記憶装置130へ更新されたデータを送信するために、記憶装置130と接続している入出力パスを制御するチャネル制御プロセッサ310へ処理を引き継ぐ。ただし、クライアント105との通信をするチャネル制御プロセッサ310がそのまま処理を続行してもよい。この場合、本処理は省略出来る（ステップ670）。
【００９０】
記憶装置130と通信を開始するため、キャッシュストレージ125はまず記憶装置130に対して、キャッシュストレージ125が使用する領域のロック要求を発行する。このときキャッシュストレージ125はReserveコマンドを使用し、自身が有するアドレス対応テーブル226を参照して指定する領域を決定する（ステップ672）。
【００９１】
Reserveコマンドを受けた記憶装置130は、Reserveコマンドで指定された記憶装置130上の領域のロック状態を確認する。
指定された領域がロックされていなければ、記憶装置130はその領域をロックし、ステータス"Good"をキャッシュストレージ125へ送信する。
指定された領域が既にロックされていれば、記憶装置130はステータス"Reservation conflict"をキャッシュストレージ125へ送信する。
このとき、ロック要求を発行したキャッシュストレージ125の状況を確認するために、記憶装置130は、ロックOKを示すコマンドをキャッシュストレージ125へ発行する（ステップ676）。
【００９２】
ロックOKを示すコマンドを受けたキャッシュストレージ125は、その確認応答のためにロックAcknowledgeを記憶装置130に送信する。これは、他のキャッシュストレージ、クライアントからロック要求を受けたときにデットロックを生じさせないための手段として用

いられる（ステップ678）。

【００９３】

ロックAcknowledgeを送信したキャッシュストレージ125は、指定した領域に対応するロック管理テーブルを更新（ここでは図７の（３）から（４）の内容へ更新）する。又、ロックAcknowledgeを受信した記憶装置130は、指定された領域に対応するロック管理テーブルを更新（ここでは、図８の（１）から（２）の内容へ更新）する（ステップ674）。

【００９４】

ロック管理テーブル225を更新したキャッシュストレージ125は、記憶装置130に対してデータ更新のI/Oリクエストを発行する（ステップ680）。

I/Oリクエストを受信した記憶装置130は、そのリクエストを処理し、ステータスを含む応答をキャッシュストレージ125へ送信する（ステップ682）。

【００９５】

その後、必要に応じてステップ680、682の処理を繰り返し、一連の処理が終了するとキャッシュストレージ125は該領域のロック解除要求を発行する。このときロック解除要求としてReleaseコマンドが使用される（ステップ684）。

Releaseコマンドを受信した記憶装置130は、Releaseコマンドで指定される領域を開放し、ステータス"Good"をキャッシュストレージ125へ送信する。

【００９６】

このとき、Reserve時と同様にロック解除要求を発行したキャッシュストレージ125の状況を確認するために、記憶装置１３０は、ロック解除OKを示すコマンドをキャッシュストレージ125に発行する（ステップ686）。

ロック解除OKを示すコマンドを受けたキャッシュストレージ125は、その確認応答のために解除Acknowledgeを記憶装置130に送信する（ステップ688）。

【００９７】

ロック解除OKを示すコマンドを送信したキャッシュストレージ125は、該領域に対応するロック管理テーブル225を更新（この場合、図７の（４）から（５）の内容へ更新）する。又、ロック解除OKを示すコマンドを受信した記憶装置130は、該領域に対応するロック管理テーブル225を更新（この場合、図８の（２）から（３）の内容へ更新）する（ステップ690）。

【００９８】

上述の処理手順により、複数のクライアント105によるアクセス競合が発生した場合でも、矛盾無くユーザデータの更新を行うことができる。

尚、上述したロック制御において、キャッシュストレージ125と記憶装置130との間の通信を暗号化することも出来る。ここで、双方の装置にキーを与える。このキーは、キャッシュストレージ125と記憶装置130間の通信を暗号化するための使用されるキーである。

【００９９】

キーは、共有秘密鍵方式、公開鍵方式などあるがここでは特に定めない。暗号化通信は図６のステップ672からステップ688までで行われる。尚、ロック制御に関わらず、クライアント105とキャッシュストレージ125、キャッシュストレージ125と記憶装置130との間の通信についても暗号化通信を採用することもできる。

【０１００】

上述の処理手順では、クライアント105とキャッシュストレージ125との間でのロック処理の後、キャッシュストレージ125と記憶装置130との間でのロック処理を独立して行っていた。しかし、別のロック制御の方法として、クライアント105が直接記憶装置130をロックする方法がある。

【０１０１】

図９は、クライアント105が記憶装置130を直接ロックする場合の手順例を示す図である。

クライアント105が記憶装置130の領域をロックするために、まずクライアント105がキャッシュストレージ125に対してロック要求を発行する（ステップ905）。

【０１０２】
ロック要求を受信したキャッシュストレージ125は、記憶装置130に対してロック要求を発行する。このとき、キャッシュストレージ125はReserveコマンドを使用し、該領域の情報をアドレス対応テーブルを参照し決定する（ステップ910）。
【０１０３】
Reserveコマンドを受信した記憶装置130は、Reserveコマンドで指定された記憶装置上の該領域のロック状態を確認する。
記憶装置130は、該領域がロックされていなければロックしステータス"Good"をキャッシュストレージ125経由でクライアント105へ送信する。
Reserveコマンドで指定された領域が既にロックされている場合、記憶装置130はステータス"Reservation conflict"をキャッシュストレージ125経由でクライアント130へ送信する。このとき、ロック要求を発行したクライアント105の状況を確認するためにロックOKを示すコマンドをキャッシュストレージ125経由で発行する（ステップ920、930）。
【０１０４】
ロックOKを示すコマンドを受けたクライアント105は、その確認応答のためにロックAcknowledgeをキャッシュストレージ125経由で記憶装置130へ送信する（ステップ935、940）。
ロックAcknowledgeを受信した記憶装置130は、該領域に対応するロック管理テーブルを更新（ここでは、図８の（１）から（２）の内容へ更新）する（ステップ925）。
【０１０５】
ロックAcknowledgeを送信したクライアント105は、キャッシュストレージ125に対してI/Oリクエストを発行する（ステップ945）。
I/Oリクエストを受信したキャッシュストレージ125は、そのリクエストを処理しステータスを含む応答をクライアント105へ送信する（ステップ950）。
【０１０６】
必要に応じてステップ945、950の処理を繰り返し、一連の処理が終了すると、キャッシュストレージ125は記憶装置130へそのI/Oリクエストに対応するデータを送信する処理を開始するが、そのタイミングはクライアント105からの一連の処理終了直後など任意のタイミングで行ってよい。
【０１０７】
キャッシュストレージ125は、クライアント105から受信したデータを記憶装置130へ送信するためのI/Oリクエストを発行する（ステップ955）。
【０１０８】
I/Oリクエストを受信した記憶装置130は、そのI/Oリクエストに対応する処理を行い、ステータスをキャッシュストレージ125へ送信する（ステップ960）。
必要に応じてステップ955、960を繰り返し、一連の処理が終了すると、キャッシュストレージ125はクライアント105に対して終了を報告する。この報告には、iSCSIのAshynchronous Messageなどが用いられる（ステップ962）。
【０１０９】
終了の報告を受けたクライアント105は、ロック解除要求をキャッシュストレージ125経由で記憶装置130へ送信する（ステップ965、970）。
ロック解除要求を受信した記憶装置130は、ロック解除要求で指定された領域を解放し、ステータスをキャッシュストレージ125経由でクライアント105へ送信する（ステップ972、974）。
【０１１０】
このとき、Reserve時と同様にロック解除要求を発行したクライアント105の状況を確認するために、記憶装置130はロック解除OKを示すコマンドをキャッシュストレージ経由でクライアント105へ発行する（ステップ972、974）。
ロック解除OKを受信したクライアント105は、その確認応答のために解除Acknowledgeをキャッシュストレージ125経由で記憶装置130に送信する（ステップ976、978）。

【０１１１】
解除Acknowledgeを受信したキャッシュストレージ125は、該領域に対応するロック管理テーブル225を更新（この場合、図７の（４）から（５）の内容へ更新）する。又、解除Acknowledgeを受信した記憶装置130は、該領域に対応するロック管理テーブル225を更新（この場合、図８の（２）から（３）の内容へ更新）する（ステップ980）。
【０１１２】
上述のロック制御によって、キャッシュストレージ125、またはキャッシュストレージ125と記憶装置130とが他からのアクセスを拒否（ロック）することによって複数のクライアント105からリクエストを受けてもデータの更新順序を補償する事が出来る。
なお、上述の処理においてキャッシュストレージ125がネットワーク1上に存在しても同じである。更に上述した暗号化通信を、図９のステップ950以降からステップ962までで行っても良い。
【０１１３】
以下、更に別のロック制御について説明する。
図１０は、クライアント105、キャッシュストレージ125及び記憶装置間の通信フローを示している。ここでは、クライアント105がキャッシュストレージ125のデバイスにログインするときに領域のロックが行われる。
【０１１４】
ログイン要求1105をクライアント105から受信したキャッシュストレージ125は、キャッシュストレージ125内のロック管理テーブル225のキャッシュロックフラグを確認する。
キャッシュストレージ125は、キャッシュロックフラグがOFFの場合はロック要求の受諾、ONの場合は拒否応答をクライアント105へ送信する（ステップ1110）。
【０１１５】
ロック要求受諾を受けたクライアント105は、キャッシュストレージ125へI/Oリクエストを発行する（ステップ1130）。
I/Oリクエストを受信したキャッシュストレージ125は、そのI/Oリクエストを処理し、そのステータスをクライアント105へ送信する（ステップ1135）。
【０１１６】
その間他のクライアント105からのログイン要求（ステップ1120）、リクエスト（ステップ1125）を受けても、キャッシュストレージ125は拒絶する。
必要に応じてステップ1130、1135の処理を繰り返し処理が終了すると、キャッシュストレージ125も該領域のロックを解除する（ステップ1140）。
【０１１７】
この後、キャッシュストレージ125は記憶装置130へ該データを送信するために記憶装置130と接続している入出力パスを制御するチャネル制御プロセッサ310へ処理を引き継ぐ。ただし、クライアントとの通信をするチャネル制御プロセッサがそのまま処理を続行してもよい。その場合、本処理は省略出来る（ステップ1145）。
【０１１８】
その後のキャッシュストレージ125と記憶装置130との通信手順は、上述したクライアント105とキャッシュストレージ125との通信手順と同様となる。
このようなロック制御をすることで、ログイン認証時に一括して処理を行うことが出来るためネットワーク上へ送出するパケット量を抑止することが出来る。
【実施例２】
【０１１９】
以下、本発明に係わるキャッシュストレージ装置の実施例２について説明する。
以下特に説明のない部分は実施例１と同じとする。
本実施形態は、図１においてネットワーク１と２との間を接続するネットワーク結合装置140がアクセス代理装置1060に置き換わる点で実施例１と異なる。
【０１２０】
アクセス代理装置1060は、クライアント105からのリクエストを代理で記憶装置130やキ

ャッシュストレージ125に対して行う。

具体的には、クライアント105からのリクエストをアクセス代理装置1060が受け取り、そのリクエストに従いアクセス代理装置1060がキャッシュストレージ125と通信を開始する。

通信終了後、ステータス等をアクセス代理装置1060がキャッシュストレージ125等から受け取り、クライアント105へ送信する。

【０１２１】

本実施例によれば、クライアント105が属するネットワーク（ネットワーク1）と記憶装置130が属するネットワーク（ネットワーク2）をアクセス代理装置1060が中継することで、ネットワーク1へ不正なデータ流入を防止しつつ、キャッシュストレージ125によるレスポンス性能向上、トランザクション性能が向上する。

【０１２２】

以下、上述の実施例のネットワークシステムにおける記憶装置130、キャッシュストレージ125及びクライアント105との間のデータの対応関係の例を幾つか説明する。

【０１２３】

ここで、これらのクライアント、キャッシュストレージ及び記憶装置の対応関係についての情報はネームサービス110に登録する。これらの情報の設定、変更などは管理端末などからネットワーク135等を介してネームサービス110に行う。

図１４は、データの対応関係の第一の例を示す図である。

【０１２４】

クライアント105a及び105bは記憶装置130と通信し、クライアント105aは記憶装置130の第1の記憶領域（一つのデバイスでもデバイスの部分でも、複数のデバイスでも構わない）１８４５を、クライアント105bは、記憶装置130の第2の記憶領域１８５０をそれぞれ使用する。尚、記憶装置130の各記憶領域は対応するクライアント専用の領域とする。すなわち、クライアント105aは第2の記憶領域にはアクセスせず、クライアント105bは第1の記憶領域にはアクセスしない。

【０１２５】

本例ではクライアント105が２つ、クライアント105と対応する記憶装置130内の記憶領域が２つであるが、クライアントの数、及び記憶装置内の記憶領域の数は特に問わない。

【０１２６】

キャッシュストレージ125は、記憶装置130と同様、各クライアント105専用の領域を有し、クライアント105aには第1キャッシュ領域（１つのデバイスでも、デバイスの部分でも、複数のデバイスでも構わない）１８３５、クライアント105bには第2キャッシュ領域を割り当てているものとする。又、各キャッシュ領域はそのクライアント105専用の領域、すなわちクライアント105aから送受信されたデータは第2キャッシュ領域には蓄積されず、クライアント105bから送受信されたデータは第1キャッシュ領域に蓄積されないものとする。

【０１２７】

本例においては、クライアント105がネームサービス110を用いて記憶装置130（実際にはキャッシュデバイス125）をディスカバリする際、クライアント105に割り当てられた記憶領域のみの情報がクライアント105に返送される。

【０１２８】

より具体的には、ネームサービス110は、要求元がクライアント1の場合第1キャッシュ領域、クライアント2の場合第2キャッシュ領域に関連する情報をクライアント105に返す。このようにすれば、確実に複数のクライアント105を、各々が使用するキャッシュストレージ125の記憶領域にアクセスさせることが可能となる。

【０１２９】

図１５は、データの対応関係の別の例を示す図である。

本例では、複数のキャッシュストレージ125に単一の記憶装置130に格納されたデータが分散されてキャッシングされている。

つまり、クライアント105aのキャッシュ領域１９４０はキャッシュストレージ125aに存在し、クライアント105bのキャッシュ領域１９４５はキャッシュストレージ125bに存在する。

【０１３０】

そして、キャッシュストレージ125aから記憶装置130の記憶領域１９５０、キャッシュストレージ125bから記憶装置130の記憶領域１９５５に関する情報をキャッシュストレージ125a、125b夫々が持つ。

本例の場合、クライアント105aから記憶装置130の情報を要求されたネームサービス110は、キャッシュストレージ125aのキャッシュ領域１９４０の情報を返す。

また、クライアント105bから記憶装置130の情報を要求されたネームサービス110は、キャッシュストレージ125bのキャッシュ領域１９４５の情報を返す。

【０１３１】

本例によれば、1つ以上のクライアント105のデータが分散されて管理されている環境下で記憶装置へそのキャッシュデータを安全に格納することが出来る。

【０１３２】

図１６は、データの対応関係の別の例を示す図である。

本例では、複数のクライアント105が記憶装置130内の同一の記憶領域にアクセスするとし、そのデータが分散されて複数のキャッシュストレージ125に格納される。

【０１３３】

ネームサービス110は、クライアント105a又は105bから記憶装置130の情報を要求されると、クライアント105aに対してはキャッシュストレージ125aのキャッシュ領域２０４０の情報を、クライアント105bに対してはキャッシュストレージ125bのキャッシュ領域２０４５の情報を返す。

【０１３４】

尚、本例では、記憶装置130の同一の記憶領域を2つのクライアント105で共有することになり互いのアクセスが衝突する可能性がある。

このアクセスの衝突を回避するために、本例では、実施例１で説明したように、各クライアント105からキャッシュストレージ125を介して記憶装置130をロックする。

【実施例３】

【０１３５】

以下、本発明に係わるキャッシュストレージ装置の実施例３を図面に示しさらに詳細に説明する。

以下特に説明のない部分は実施例１及び２と同じとする。

本実施例では、実施例１及び２では独立して存在していたネームサービスをキャッシュストレージ125内に組み込んだ点がこれまでの実施例とは異なる。

【０１３６】

図１７は、本実施例のキャッシュストレージ125を含んだネットワークシステムの例を示す図である。本ネットワークシステムは、クライアント105、キャッシュストレージ125及び記憶装置130がネットワーク２１０５を介して相互に接続されている。

【０１３７】

本実施例のキャッシュストレージ125は、実施例1、２で示した複数のチャネル制御部２１０をプロトコル処理部２１２０、キャッシュメモリ制御部２１５、共有メモリ制御部２２０を総称してメモリ制御部２１２５、複数のディスク制御部２３０をI/O処理部２１３０で示している。さらに、キャッシュストレージ１２５のプロトコル処理部２１２０にはネームサービス提供部２１４５を備える。

【０１３８】

本実施例の記憶装置130は、本実施例のキャッシュストレージ125のネームサービス提供部2145を除いた構成に等しい。

プロトコル処理部２１２０とメモリ制御部２１２５、メモリ制御部２１２５とI/O処理部２１３０．ディスク駆動装置235はそれぞれ相互に接続される。

【０１３９】
プロトコル処理部２１２０は1つ以上の通信ポートを有し、各通信ポートにはネットワーク識別子が割り振られている。
【０１４０】
ネットワーク識別子とは各通信ポートを識別するための情報で例えばIPアドレス、MACアドレス、ポート番号である。
【０１４１】
ネームサービス提供部２１４５は、ネットワーク識別子、キャッシュストレージ125のディスク駆動装置235のディスクドライブに割り当てたiSCSI Name、記憶装置130のネットワーク識別子、iSCSI Name、ネットワーク２１０５上のその他のデバイス情報を管理し、クライアントなどデバイスからの問い合わせに対し該当する情報を通知するための情報を保持している。
【０１４２】
なお、実施例３ではプロトコル処理部２１２０、結合部２１２５、Ｉ／Ｏ処理部２１３０、ネームサービス提供部２１４５、ディスク駆動装置２３５は同一筐体内に実装しているが、別筐体で実装しても同等の機能を実現することが出来る。
【０１４３】
本実施例では、実施例１や２で説明したクライアント105のディスカバリリクエストが、キャッシュストレージ125に送信される。このリクエストを受信したキャッシュストレージ125は、クライアント105がアクセスしようとしている記憶装置130の代わりに、自分自身の情報をクライアント105に送信する。後の処理は、実施例１又は２と同様である。
【０１４４】
尚、同一のネットワーク内にキャッシュストレージ125が複数ある場合、どちらか一方がネームサービスの処理を行っても、双方で分担して処理を行っても良い。双方で分担して処理を行う場合、複数のクライアント105の各々はどちらかのネームサービスを有するキャッシュストレージ125へディスカバリリクエストを送信するように予め決めておけばよい。
【０１４５】
実施例３によれば、ネームサービスを記憶装置内に実装することで外部にネームサービスへの問い合わせが不要になる、かつネームサービスが管理している情報の機密性が向上する。
【図面の簡単な説明】
【０１４６】
【図１】ネットワークシステムの全体構成例を示す図である。
【図２】記憶装置及びキャッシュストレージの構成例を示す図である。
【図３】チャネル制御部の構成例を示す図である。
【図４】ロック管理テーブルの例を示す図である。
【図５】クライアントから記憶装置間の通信フローの例を示す図である。
【図６】クライアントから記憶装置間の通信フローの例を示す図である。
【図７】キャッシュストレージにおけるロック管理テーブルの内容の変化の例を示す図である。
【図８】記憶装置におけるロック管理テーブルの内容の変化の例を示す図である。
【図９】クライアントから記憶装置間の通信フローの例を示す図である。
【図１０】クライアントから記憶装置間の通信フローの例を示す図である。
【図１１】アドレス対応テーブルの例を示す図である。
【図１２】伝送路で障害が発生した場合の処理例を示す図である。
【図１３】ネットワークパケットの例を示す図である。
【図１４】クライアント、キャッシュストレージ及び記憶装置の間のデータの関係例を示す模式図である。
【図１５】クライアント、キャッシュストレージ及び記憶装置の間のデータの関係例

を示す模式図である。

【図１６】クライアント、キャッシュストレージ及び記憶装置の間のデータの関係例を示す模式図である。

【図１７】記憶装置の構成例を示す図である。

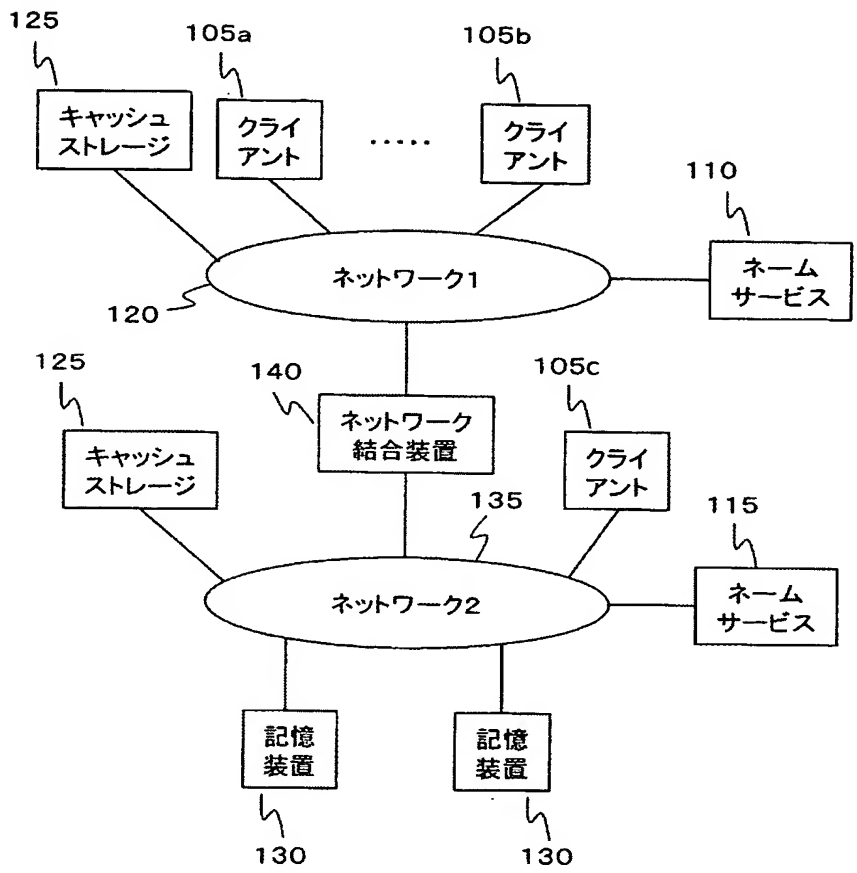
【符号の説明】

【０１４７】

２０６…入出力パス、２１０…チャネル制御部、２１５…チャネル制御部、２２０…共有メモリ制御部、２２５…ロック管理テーブル、２２６…アドレス対応テーブル、２３０…ディスク制御部、２３５…ディスク制御装置、２４０…ディスクドライブ。

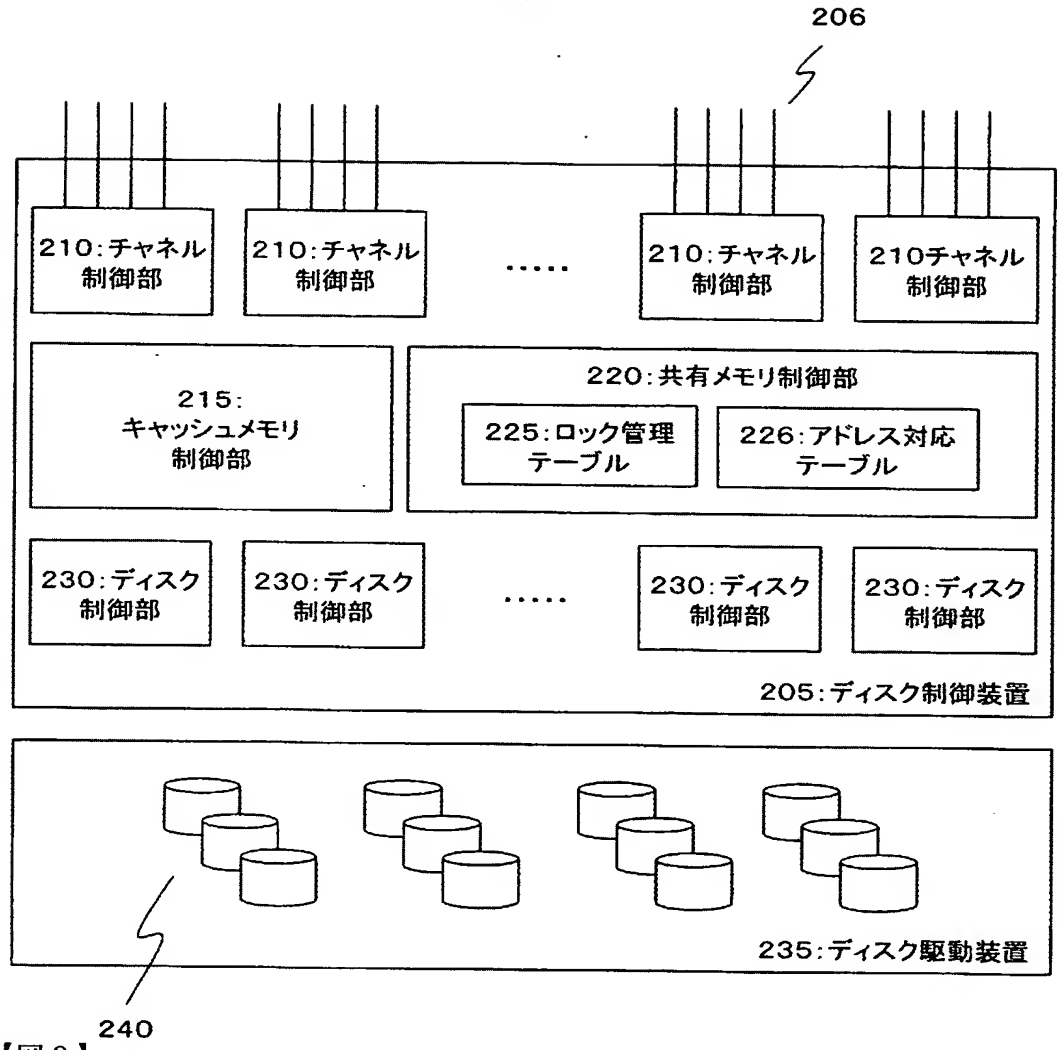
【書類名】図面
【図１】

図1


125
105a
105b
キャッシュストレージ
クライアント
.....
クライアント
110
ネットワーク1
ネームサービス
120
125
140
105c
キャッシュストレージ
ネットワーク結合装置
クライアント
135
115
ネットワーク2
ネームサービス
記憶装置
記憶装置
130
130

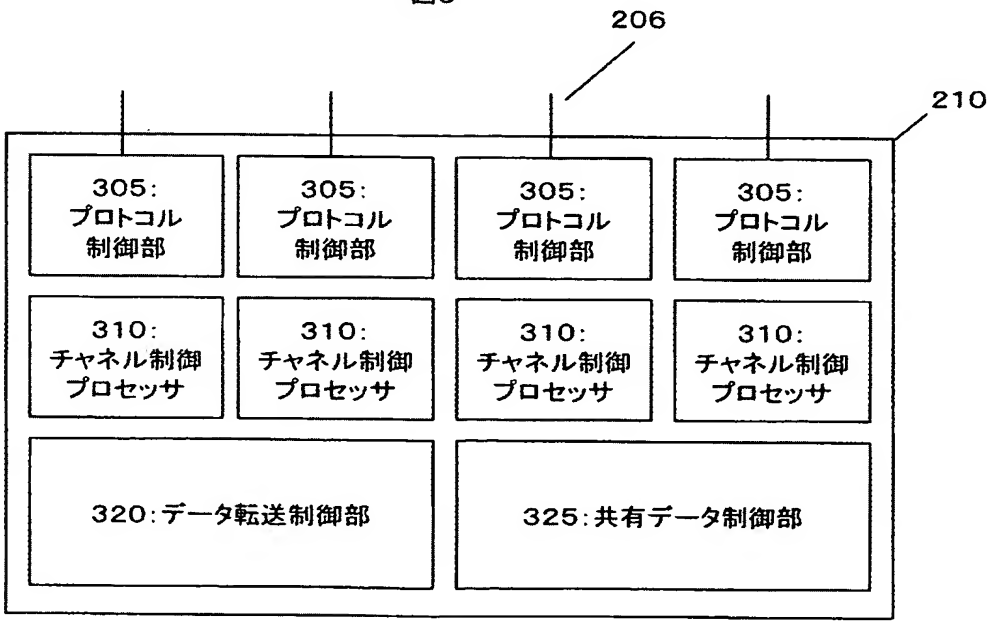
【図２】

図2


206
210:チャネル制御部
210:チャネル制御部
.....
210:チャネル制御部
210チャネル制御部
215:キャッシュメモリ制御部
220:共有メモリ制御部
225:ロック管理テーブル
226:アドレス対応テーブル
230:ディスク制御部
230:ディスク制御部
.....
230:ディスク制御部
230:ディスク制御部
205:ディスク制御装置
235:ディスク駆動装置
240

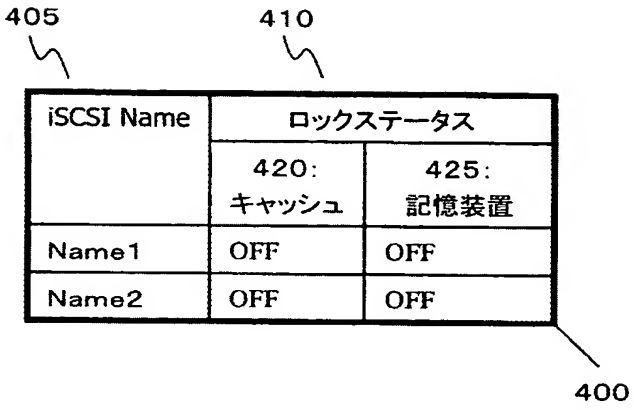
【図３】

図3


206
210
305:プロトコル制御部
305:プロトコル制御部
305:プロトコル制御部
305:プロトコル制御部
310:チャネル制御プロセッサ
310:チャネル制御プロセッサ
310:チャネル制御プロセッサ
310:チャネル制御プロセッサ
320:データ転送制御部
325:共有データ制御部

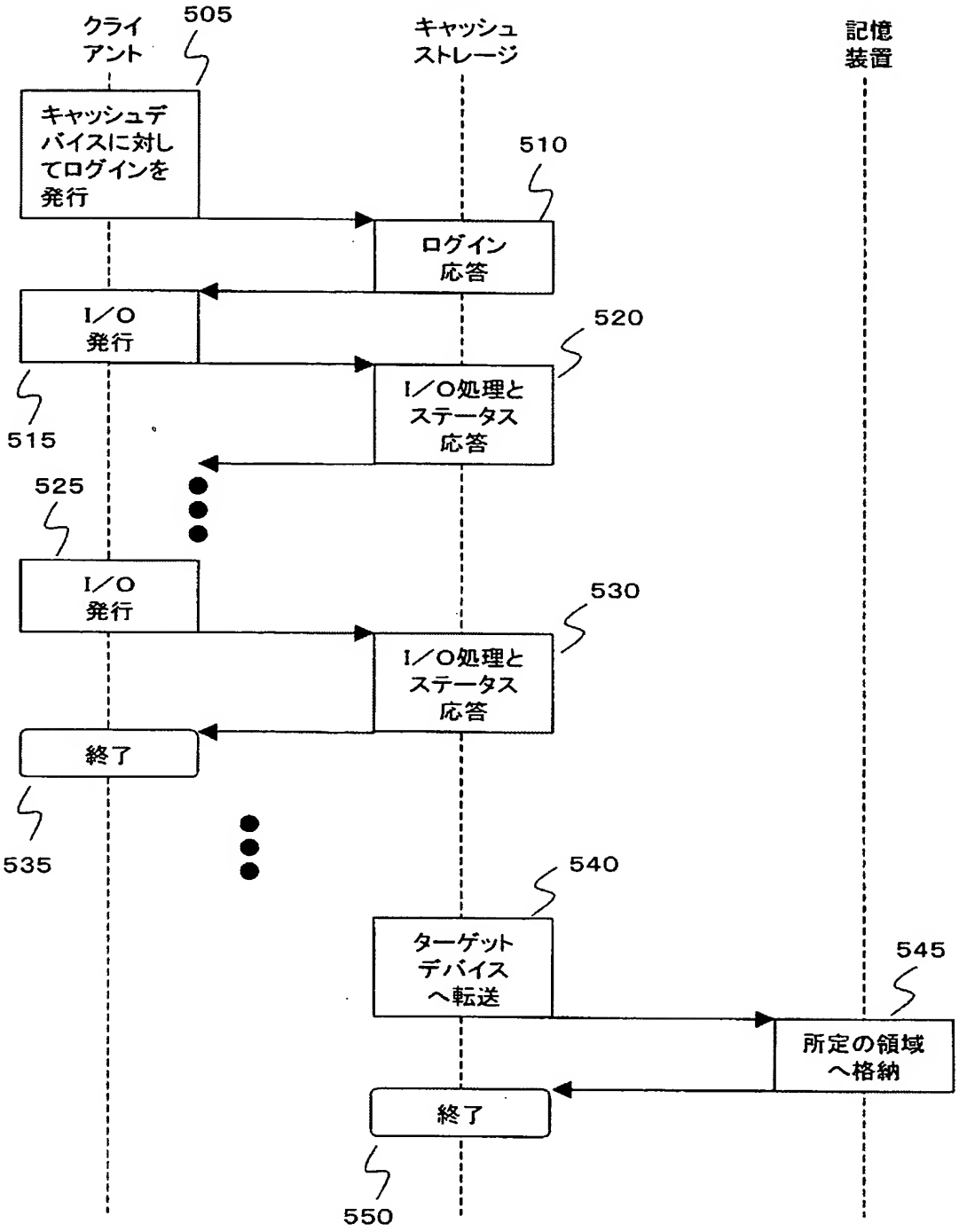
【図４】

図4


405
410
iSCSI Name
ロックステータス
420:
キャッシュ
425:
記憶装置
Name1
OFF
OFF
Name2
OFF
OFF
400

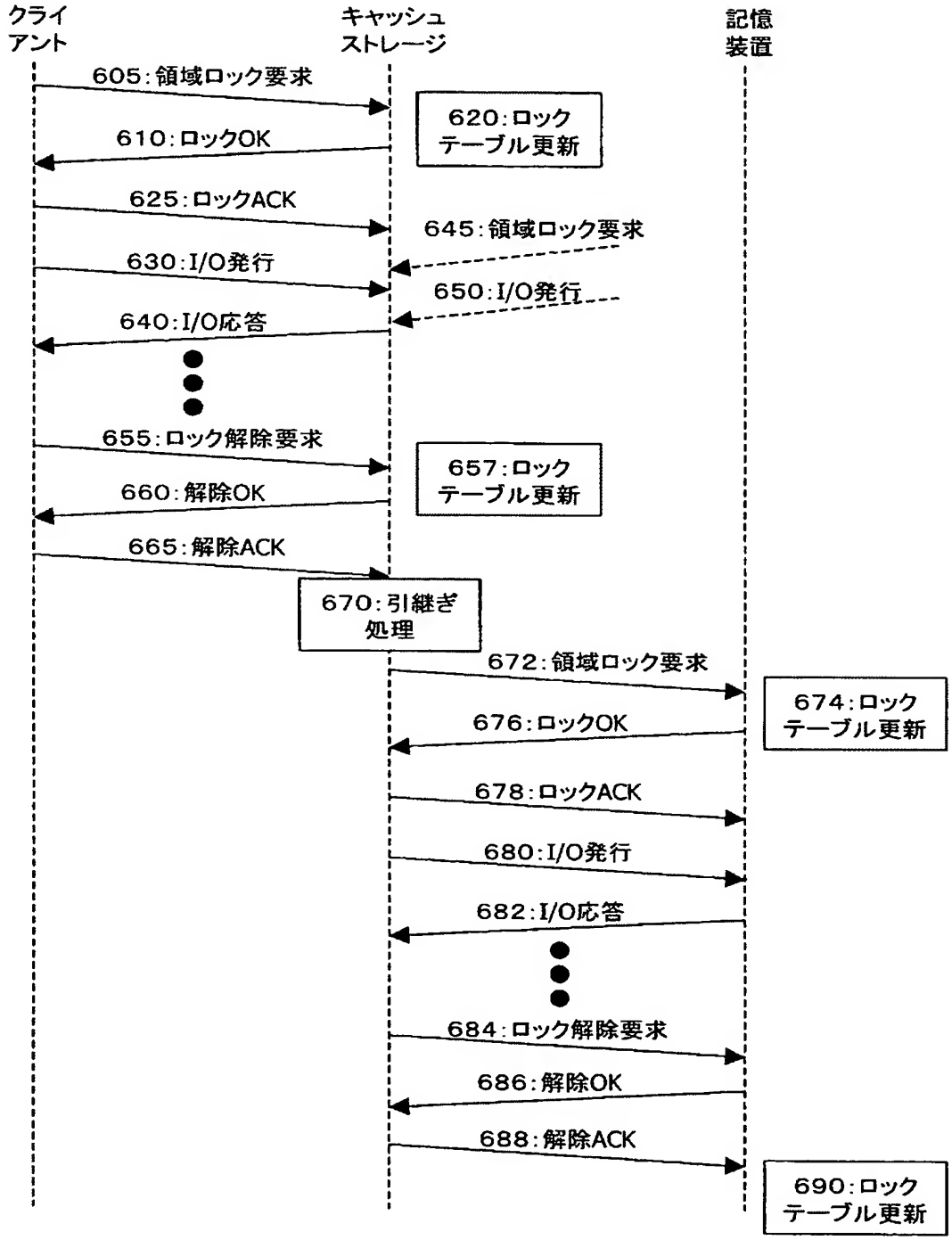
【図５】

図5


クライアント
505
キャッシュストレージ
記憶装置
キャッシュデバイスに対してログインを発行
510
ログイン応答
I／O発行
520
I／O処理とステータス応答
515
525
I／O発行
530
I／O処理とステータス応答
終了
535
540
ターゲットデバイスへ転送
545
所定の領域へ格納
終了
550

【図６】

図6


クライアント
キャッシュストレージ
記憶装置
605:領域ロック要求
620:ロックテーブル更新
610:ロックOK
625:ロックACK
645:領域ロック要求
630:I/O発行
650:I/O発行
640:I/O応答
655:ロック解除要求
657:ロックテーブル更新
660:解除OK
665:解除ACK
670:引継ぎ処理
672:領域ロック要求
674:ロックテーブル更新
676:ロックOK
678:ロックACK
680:I/O発行
682:I/O応答
684:ロック解除要求
686:解除OK
688:解除ACK
690:ロックテーブル更新

【図７】

図7

(1)

Index	iSCSI Name	ロックステータス	
		キャッシュ	ストレージ
dev1	Name1	OFF	OFF
dev2	Name2	OFF	OFF

(2)

Index	iSCSI Name	ロックステータス	
		キャッシュ	ストレージ
dev1	Name1	ON	OFF
dev2	Name2	OFF	OFF

(3)

Index	iSCSI Name	ロックステータス	
		キャッシュ	ストレージ
dev1	Name1	OFF	OFF
dev2	Name2	OFF	OFF

(4)

Index	iSCSI Name	ロックステータス	
		キャッシュ	ストレージ
dev1	Name1	OFF	ON
dev2	Name2	OFF	OFF

(5)

Index	iSCSI Name	ロックステータス	
		キャッシュ	ストレージ
dev1	Name1	OFF	OFF
dev2	Name2	OFF	OFF

【図８】

図8

(1)

Index	iSCSI Name	ロックステータス	
		キャッシュ	ストレージ
dev1	Name1	OFF	OFF
dev2	Name2	OFF	OFF

(2)

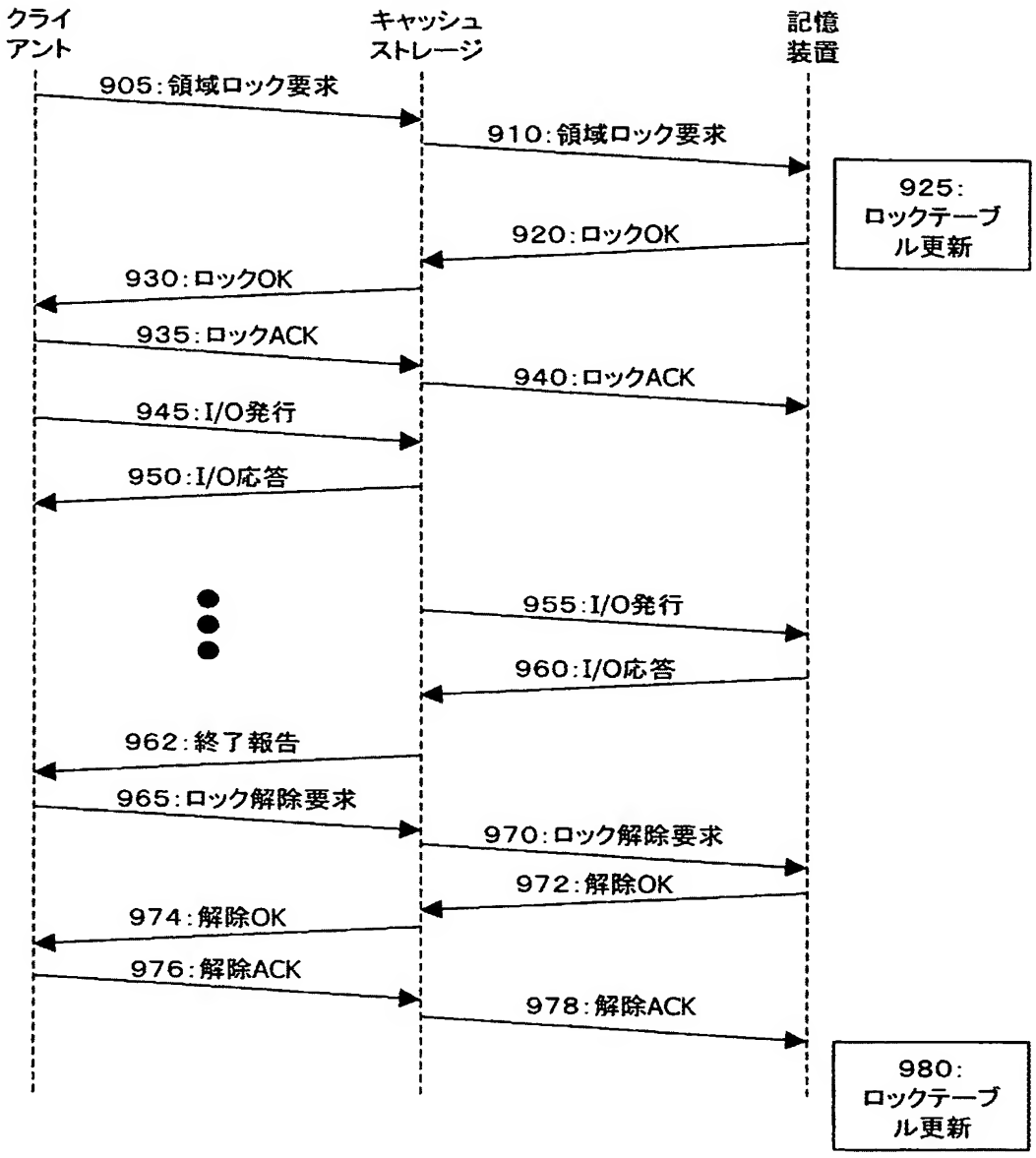
Index	iSCSI Name	ロックステータス	
		キャッシュ	ストレージ
dev1	Name1	OFF	ON
dev2	Name2	OFF	OFF

(3)

Index	iSCSI Name	ロックステータス	
		キャッシュ	ストレージ
dev1	Name1	OFF	OFF
dev2	Name2	OFF	OFF

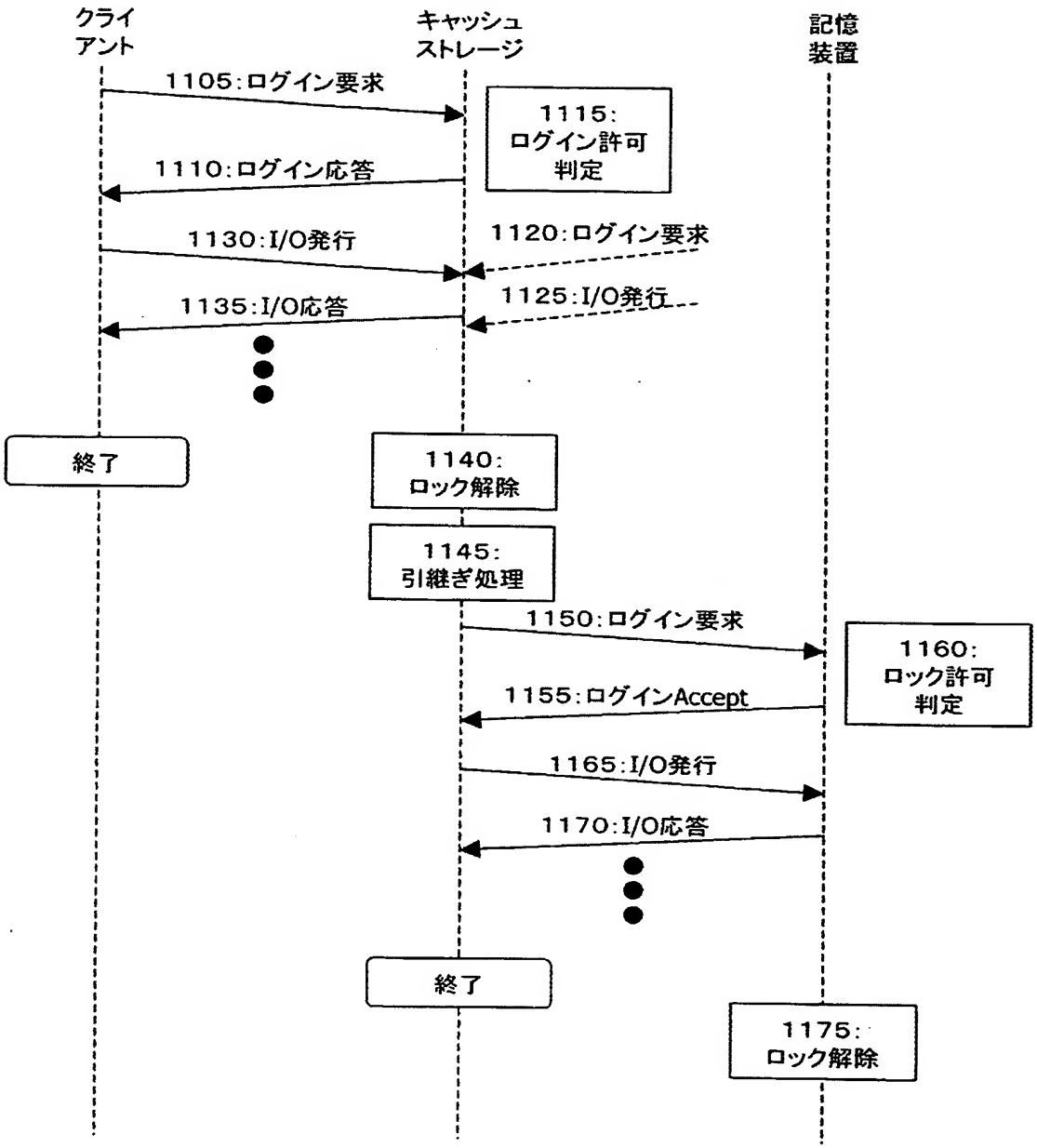
【図９】

図9


クライアント
キャッシュストレージ
記憶装置
905:領域ロック要求
910:領域ロック要求
925:ロックテーブル更新
920:ロックOK
930:ロックOK
935:ロックACK
940:ロックACK
945:I/O発行
950:I/O応答
955:I/O発行
960:I/O応答
962:終了報告
965:ロック解除要求
970:ロック解除要求
972:解除OK
974:解除OK
976:解除ACK
978:解除ACK
980:ロックテーブル更新

【図１０】

図10



【図１１】

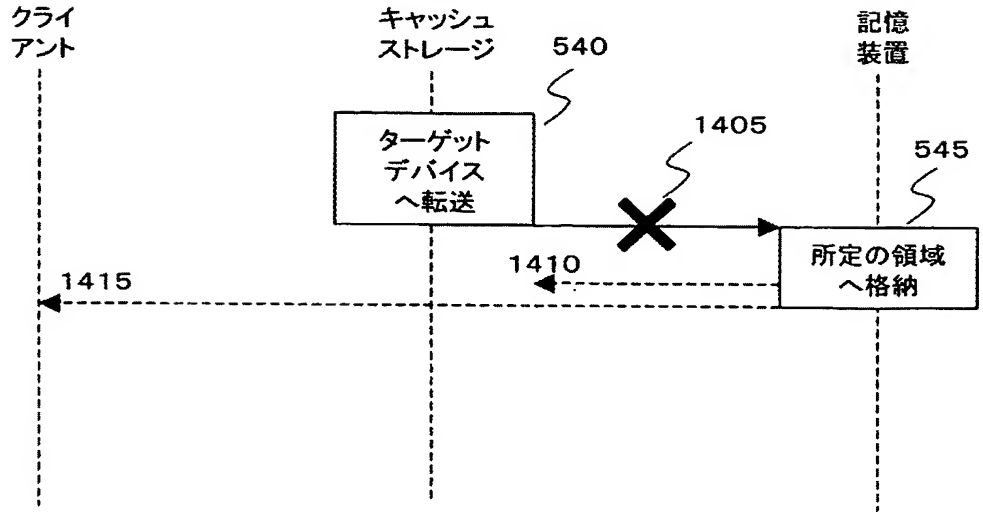
図11

1305:デバイス名	1310:アドレス
Dev1	Name1
Dev2	Name2
Dev3	Name3

1300

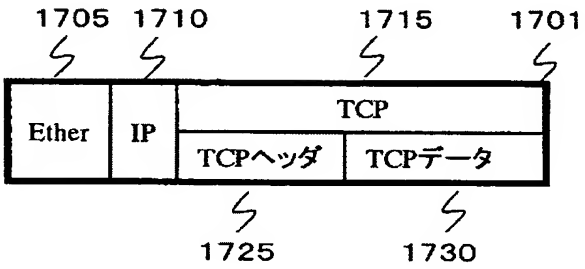
【図１２】

図12



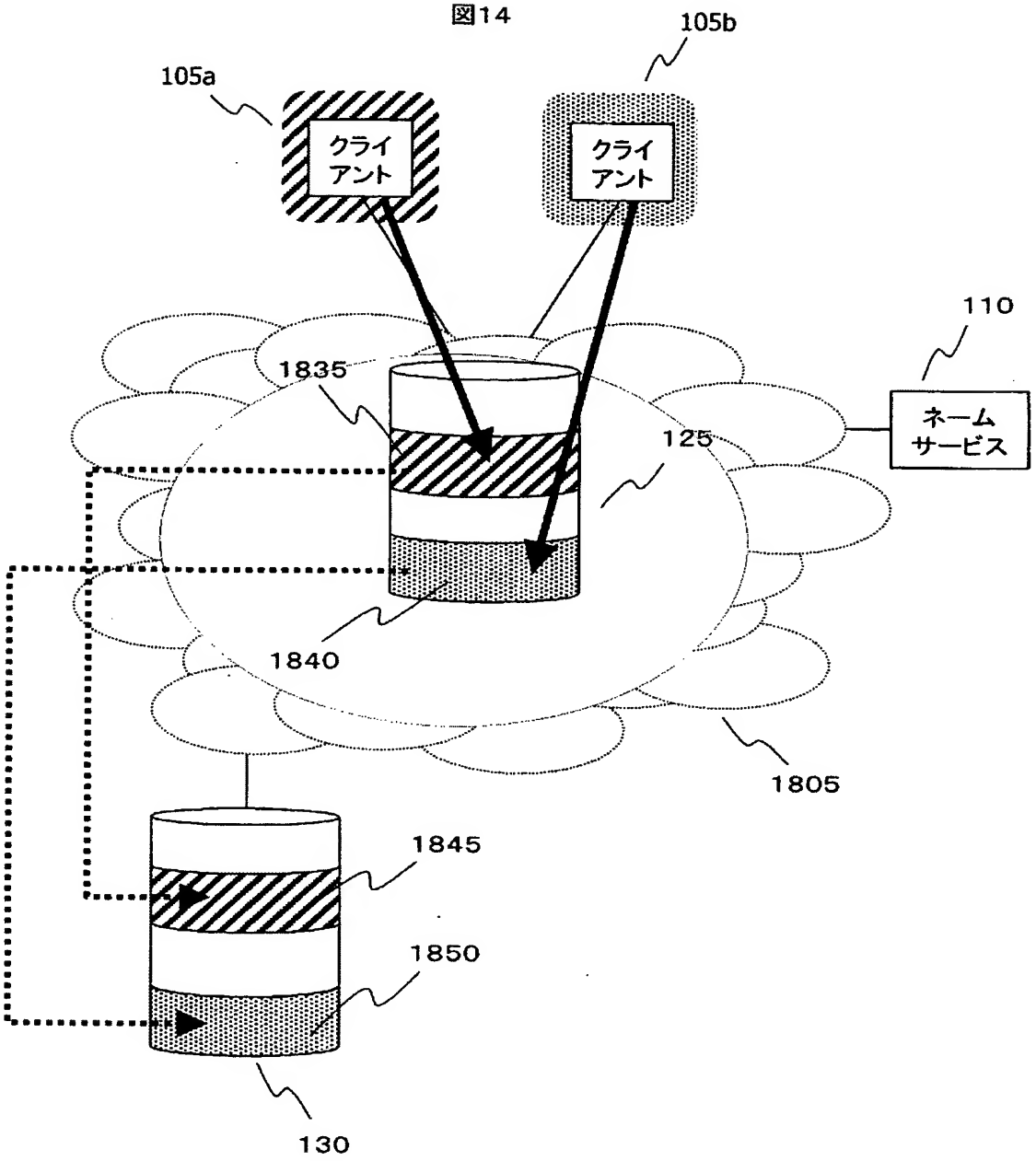
【図１３】

図13



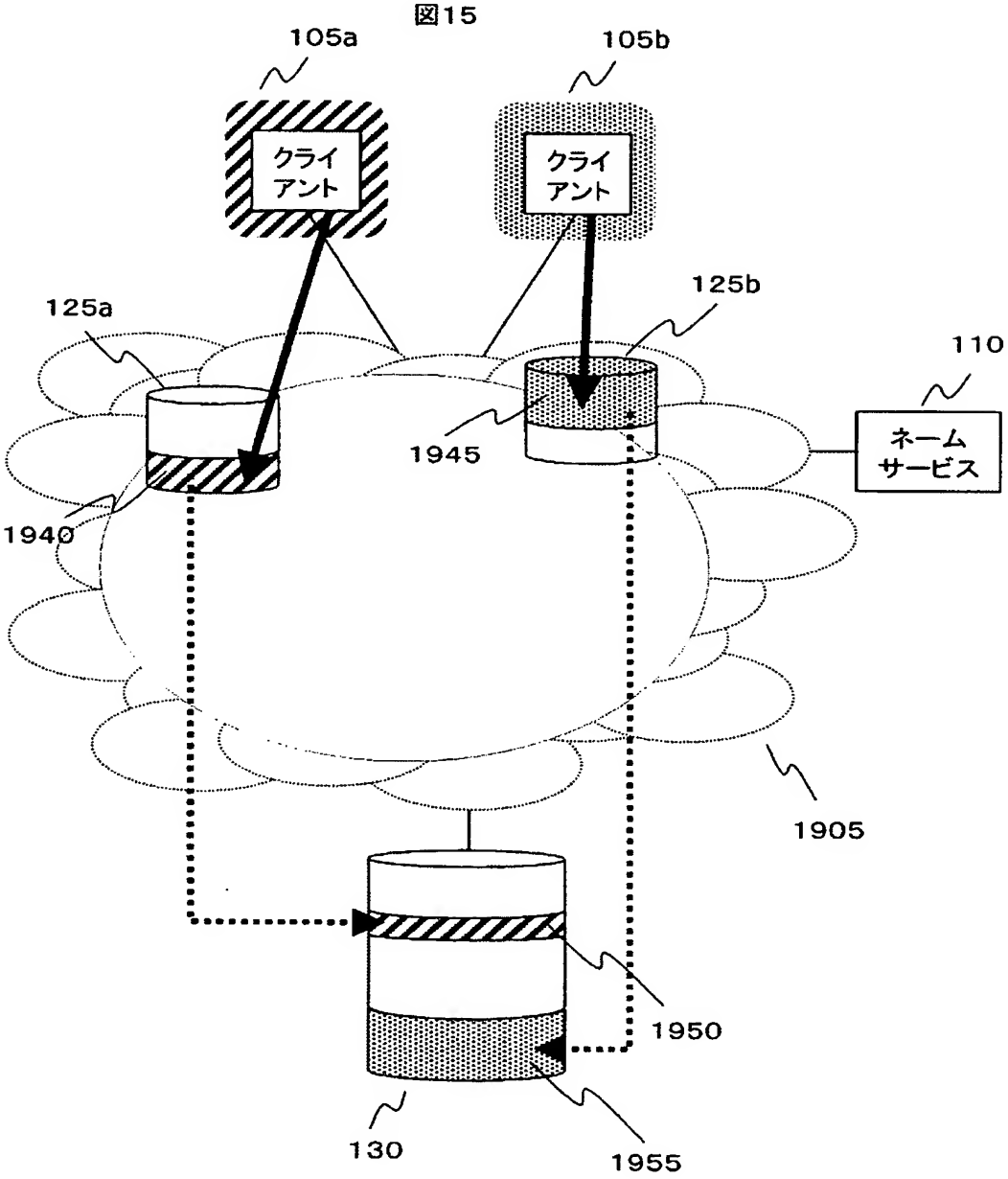
【図１４】

図14


105a
105b
クライアント
クライアント
110
1835
ネームサービス
125
1840
1805
1845
1850
130

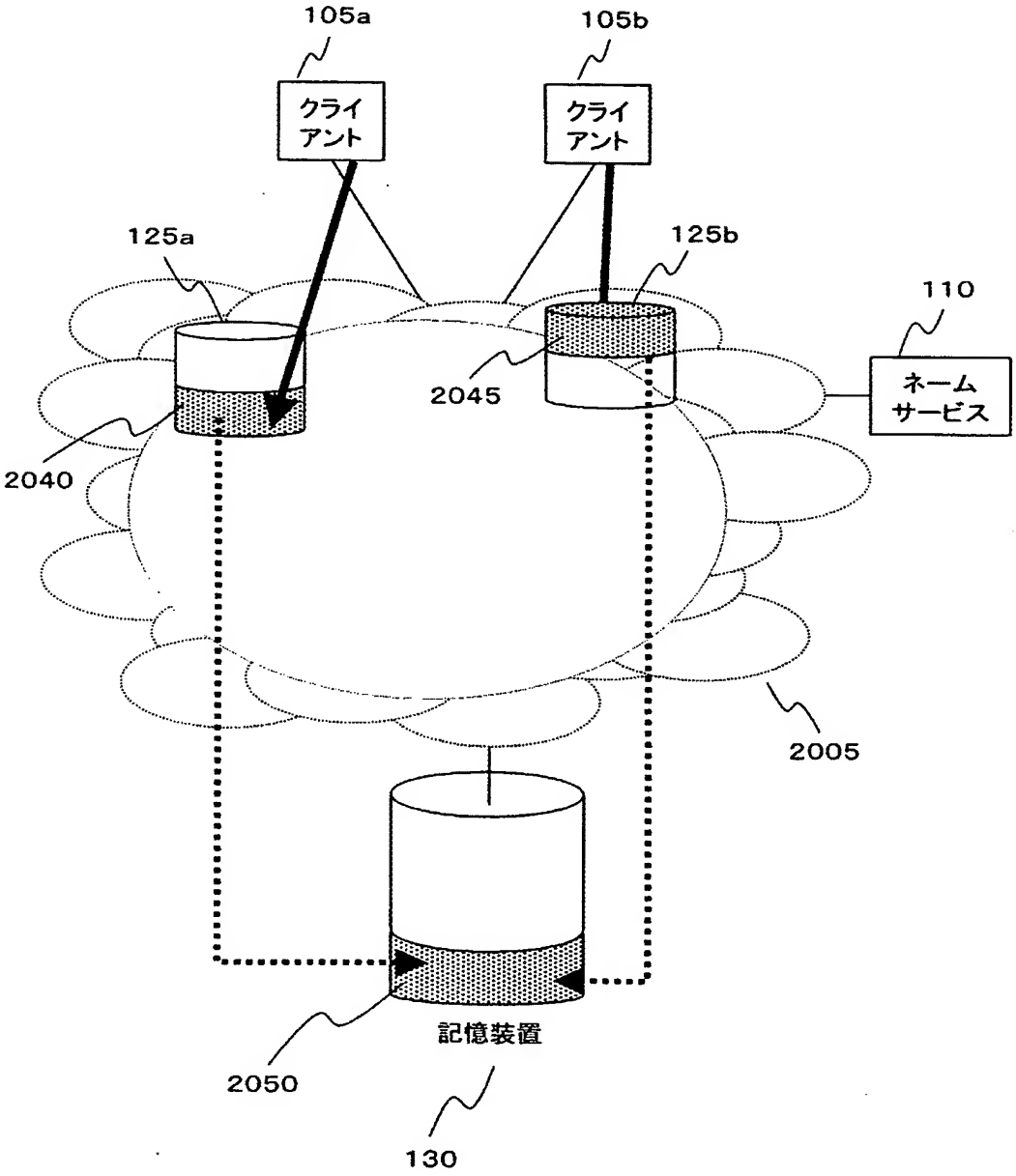
【図１５】

図15


105a
105b
クライアント
クライアント
125a
125b
110
ネームサービス
1945
1940
1905
1950
130
1955

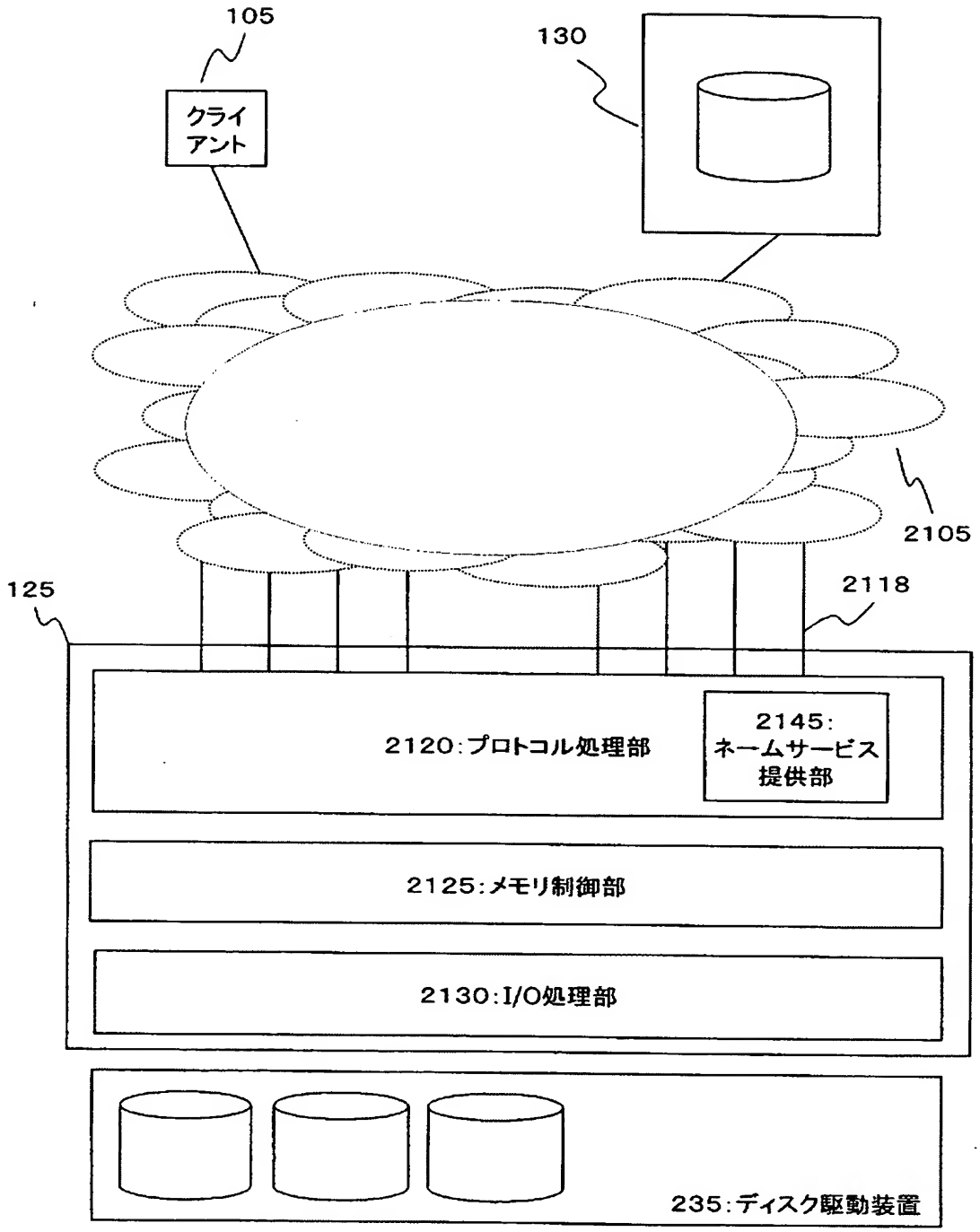
【図１６】

図16


105a
105b
クライアント
クライアント
125a
125b
110
ネームサービス
2045
2040
2005
記憶装置
2050
130

【図１７】

図17


105
クライアント
130
2105
125
2118
2120:プロトコル処理部
2145:
ネームサービス
提供部
2125:メモリ制御部
2130:I/O処理部
235:ディスク駆動装置

【書類名】要約書
【要約】
【課題】
TCP/IPネットワークを介してクライアントとストレージが通信する場合、伝送遅延、伝送路上の障害による再送処理で応答性能、トランザクション性能が劣化する。
【解決手段】
クライアントと記憶装置間にキャッシュストレージを設け、クライアントがアクセスする領域をあらかじめロック（排他）する。
【選択図】　図１

認定・付加情報

特許出願の番号	特願２００３－３６９８１１
受付番号	５０３０１７９７４７３
書類名	特許願
担当官	第七担当上席　０００９６
作成日	平成１５年１１月　５日

＜認定情報・付加情報＞

【提出日】　平成15年10月30日

特願２００３－３６９８１１

出　願　人　履　歴　情　報

識別番号　　［０００００５１０８］

１．変更年月日　　１９９０年　８月３１日
［変更理由］　　新規登録
住　所　　東京都千代田区神田駿河台４丁目６番地
氏　名　　株式会社日立製作所